The Americanization of Migrants' Names and its Economic Payoff

Costanza Biavaschi* Corrado Giulietti† Zahra Siddique‡

October 29, 2013

PRELIMINARY AND INCOMPLETE

Please do not cite or circulate without authors' permission

Abstract

We examine the impact of name Americanization on labor market outcomes of migrants by constructing a novel longitudinal data set of immigrant naturalization records in 1930 New York where the full initial sample is observed over time. We find that changing into an American name improves labor market outcomes. Compared to name keepers, migrants who change into more popular American names experience larger occupational upgrading, with gains of up to 14% for changing into very common names such as John or William. We show that our estimates of the returns from name Americanization are causal effects by applying instrumental variable methods where we use an index of linguistic complexity based on Scrabble points to predict changes in name Americanization. Our results highlight the tradeoffs between individual identity and labor market success, suggesting that the assimilation process in the first half of the twentieth century passed through economic as well as cultural assimilation.

JEL codes: J0, Z1, J31

Keywords: Name, Assimilation, Identity, Immigrants

*Institute for the Study of Labor (IZA). E-mail: biavaschi@iza.org

†Institute for the Study of Labor (IZA). E-mail: giulietti@iza.org

‡University of Reading. E-mail: z.siddique@reading.ac.uk

1 Introduction

And while you bring all countries with you, you come with a purpose of leaving all other countries behind you - bringing what is best of their spirit, but not looking over your shoulders and seeking to perpetuate what you intended to leave in them. [...] You cannot dedicate yourself to America unless you become in every respect and with every purpose of your will thorough Americans.

Woodrow Wilson, May 10, 1915[1]

Americanization, the process through which immigrants strive to assimilate into the American society, encompasses several dimensions. One of these is the Americanization of migrants' names, a key aspect of the desire – or need – to conform to the American norm. In this paper we provide the first evidence on the economic consequences of name Americanization at the making of modern America.

Many Americans have heard stories of migrant ancestors Americanizing their names in the early half of the twentieth century. However, such anecdotes are stored in familial memories only since, as of today, there is virtually no study measuring the extent and implications of name Americanization. Table 1 provides the first numbers on the magnitude of this phenomenon. Name Americanization was a widespread practice, with nearly a third of naturalizing immigrants abandoning their first names by 1930.[2] Panel A shows also substantial variation by country of birth. Migrants from Italy, Russia, Germany were all very likely to abandon their foreign-sounding names. These migrants were adopting names popular among the U.S. born population such as William, John, or Charles (Panel B).

INSERT TABLE 1 HERE

Widespread name Americanization brings up the natural question of whether it affected the economic success of immigrants. Figure 1 provides a preliminary answer to this question. Name Americanization into the top popular names - e.g., top quartile as we measure it - was associated with an occupation-based earnings increase of above 10%. These gains were larger than those experienced by migrants Americanizing into less popular first names - e.g., the first quartile - and even more so than those experienced by migrants who kept their name or changed to a more distinctive one.

INSERT FIGURE 1 HERE

[1]Speech given to a group of naturalized Americans in Philadelphia, retrieved from `http://www.apstudent.com/ushistory/docs1901/amrcnism.htm`

[2]Surname Americanization – as we define it in the next section – was in place but appears to be less common, with only 7% of migrants Americanizing their surnames. Details on surnames changes are given in the online Appendix. The rest of the paper focuses on first names.

We devote much of our analysis to verifying that the positive relationship between name Americanization and occupation-based earnings growth represents a credible causal effect and is not driven by observable differences across migrants. We control for characteristics which are often unavailable even in modern data. This allows us to control for time varying socio-demographic characteristics, nationality-specific and local labor market-specific time trends, obtained by geocoding of migrant addresses. More importantly, we propose to explore the causal effect of changing name by both exploiting the longitudinal nature of our data and using an instrumental variable technique. We instrument name Americanization with an index based on Scrabble points, capturing the degree of names' linguistic complexity at arrival with respect to the linguistic complexity at destination.

To understand the link between name Americanization and labor market outcomes we have digitalized a novel data set in which the same migrants are observed at different points in time. The key innovation of these data is that use a random sample of naturalization petitions filed in the Eastern and Southern Courts of New York in 1930. This differs from the common methodology used in constructing historical dataset, which, resorting to record linkage, delivers partial matching and impedes detection of name changes. Hence, our advantage is that we are able to follow the entire sample of migrants over time due to the nature of the naturalization process and documentation procedure, which required migrants to first file a declaration and later a petition for naturalization. We exploit this two-step procedure to obtain information on name Americanization and on a wide range of migrant characteristics at two different points in time, allowing us to study economic outcomes following name Americanization.

Throughout our analysis, we corroborate the persistence of a substantial return to name Americanization. Our results suggests that this practice was more common among those migrants with worse occupational trajectories. We interpret this finding with name Americanization being a way to circumvent negative occupational shocks and compensate for scarce or constrained access to other means of occupational upward mobility, such as internal mobility, human capital investments or living in ethnic enclaves.

More generally, our results highlight the tradeoff between maintaining individual identity and labor market success, suggesting that the process of migrants' advancement at the dawn of the modern melting pot involved economic as well as cultural assimilation. This view builds upon the recent evidence provided in Abramitzky et al. (2012a), which concludes that migrant occupational upgrading has been limited between 1900 and 1920. We show that name Americanization could have served as a strategy adopted by migrants to climb up the occupational ladder and that, if more had Americanized their names, more would have reached higher economic success.

We are therefore able to add evidence to the growing literature that links names to economic success for a period of time that has been crucial to the foundations of the American society. Much of the literature on the relationship between names and outcomes primarily focuses on the second half of the 1900s. For example, psychologists have shown that first names closer to host societies are associated with positive attitudes by host populations (Kang, 1971, Drury and McCarthy, 1980) and by employers, co-workers and customers (Laham et al., 2011). Economists have shown that more common names result in better educational (Figlio, 2005) and labor market success, due to a reduction in discrimination (e.g. Bertrand and Mullainathan, 2004).[3] The association between names and outcomes, however, might also stem from unobserved factors that correlate with parents naming choices and economic success. For example, Fryer and Levitt (2004) find that after controlling for background characteristics White names are not associated with better outcomes.

To our knowledge, only two studies have looked at name changes, with a focus on renouncing surnames. The first analysis looking at individual name choices is Goldin and Shim (2004).[4] The authors relate names to gender identity, focusing on women who decide to keep their surnames instead of acquiring the husband's surname upon marriage. While their work analyzes the surname choice and not on its effects on outcomes, it provides compelling evidence on the reversion of the custom of taking husband's name in the second half of the 1900s. More importantly, it documents the existence of a negative correlation between surname change and skills. The second work looks at the effects of surname changes made by immigrants from Asian, and Slavic countries living in Sweden in the 1990s (Arai and Thoursie, 2009). In this context, in which 0.4% of migrants changed surname, fixed-effects estimates show a substantial increase in earnings. This result seems driven by reduced labor market discrimination associated with a Swedish-sounding surname. This scarce number of investigations have been eluded because hardly any existing data sets contain enough information to empirically test whether changing names improves economic outcomes. Furthermore, in an historical setting like we have nailing down the consequences of name Americanization is particularly challenging because of the unavailability of longitudinal sources and the difficulties in tackling the endogenous choice of name. Our data and estimation

[3]For instance, racial or ethnic names are frequently used in audit and correspondence studies to test for racial discrimination in hiring. Bertrand and Mullainathan (2004) use a correspondence study to provide evidence that résumés with African-American names had significantly lower callback for recruitment by hiring firms than did resumes with white names. Similar studies in Europe and developing countries use ethnic first names to find evidence of labor market discrimination against immigrants, ethnic minorities and/or socio-economically disadvantaged groups.

[4]The only precedent on a similar topic is, to the best of our knowledge, the sociological study by Broom et al. (1955). The author looks at the characteristics of 1,017 petitions for name changes filed at the Los Angeles Superior Court, in 1946-1947. This study is however very descriptive in nature, it is unclear how the petitions have been selected, does not focus on migrants and does not look at labor market outcomes.

techniques overcome these limitations and contribute to the understanding of the pressures faced to conform to the American norms in the making of modern America.

Our paper is organized as follows. We first provide a background to understanding returns and tradeoffs related to name Americanization. Next we describe our data and empirical model. We then present and discuss the estimated effects of name Americanization before concluding.

2 Data

To understand the magnitude and consequences of name Americanization, this needs to be measured and defined appropriately. The ideal dataset would consist of a panel of individuals who are observed before and after name Americanization. Historical data sources, such as the censuses, do not have identifiers that allow tracking individuals over time. Hence, much of the contributions in the literature (Ferrie, 1996, Abramitzky et al., 2012a) have resorted to linking records, often matching individuals by name, age and birthplace. Name changes is in fact one of the reason for a failure of matching individuals – especially immigrants – over time, e.g. across censuses. Hence, this strategy is clearly not an option to us, as we would not be able to match individuals who have changed names.

Instead, we exploit the rich records stemming from the naturalization procedure.[5] With only marginal changes since 1795, the naturalization process consisted of two phases. Free white aliens residing in the U.S. for at least two years, one year of which was spent in the state or territory in which the application was made, were required to first file a declaration of intention. The second step involved filing a petition for admission to citizenship which could be done a minimum of five years following the initial declaration. At the moment of filing the citizenship papers, the Bureau of Immigration and Naturalization checked ship manifests and issued a certificate of arrival, a document which contains the name at entry. Finally, migrants took a naturalization oath or oath of allegiance. After a migrant had completed all citizenship requirements they were issued a certificate of naturalization. After 1906, the declaration of intention, the petition for naturalization, and the certificate of naturalization had a standardized format prescribed by the Bureau of Immigration and Naturalization, creating the rich records that we have digitalized as part of our research.

We extract the naturalization records from Ancestry.com, a website providing genealogical and family history records. We accessed digitalized files of the complete naturalization

[5] Naturalization was considered to be the ultimate act of identifying with the American culture. In fact, before the 1924 restrictions - but also after the imposition of quotas - direct benefits for citizenship were per se rather limited. Immigrants during this time were permanent residents and, before the New Deal, social benefits were too small to reflect a motivation for naturalization (Lleras-Muney and Shertzer, 2012).

records filed at the U.S. District Court for the Southern and Eastern districts of New York City in 1930.[6] More than 30% of all certificates granted in the United States pertain to the district of New York, highlighting the importance of these records in understanding immigrant assimilation in the U.S. (Annual Report of the Commissioner of Naturalization, 1930, p.15). Ancestry.com provides access to 26,113 of the official 30,361 petitions that were filed in the naturalization district of New York (Annual Report of the Commissioner of Naturalization, 1930, p.15), corresponding to more than 85% of all the records thought to be available for that year. The remaining records are likely to have been granted by the other district courts of New York City, or other courts in the State, but are not available in electronic form. Yet, the overwhelming share of naturalizations granted by the Eastern and Southern courts is an indication of the importance of these courts, making these records representative of nearly the entire population of naturalizing immigrants residing in the City.

We selected a 20% random sample of the available records for 1930 and manually transcribed them. While our randomization procedure involved the collection of naturalization records for both males and females, in the paper we use the records for male immigrants only.[7] We also restrict the sample to individuals who filed petition papers two to seven years after declaration, which gives us an estimation sample of 3,353 migrants.[8]

Given the two steps involved in the naturalization process, we can observe migrants' characteristics at two different points in time without resorting to record linkage. Besides being obvious by simple observation of the individual characteristics, we are confident that the same migrant is observed for two reasons: first, except in rare cases, documents are ordered by petition number and sequenced by document type for each migrant; and second, the petition number is usually printed in all documents.[9] In the paper, we refer to these two pieces of information as "declaration" and "petition", respectively. In order to test our main hypothesis – whether name Americanization influenced migrants' outcomes – we derived variables for both time periods. In the following paragraphs, we describe the steps we took

[6] The Southern District Court jurisdiction comprises the counties of New York (Manhattan), Bronx, Westchester, Putnam, Rockland, Orange, Dutchess, and Sullivan. The Eastern District court's territorial jurisdiction includes the counties of Kings (Brooklyn), Queens, Richmond (Staten Island), Nassau, and Suffolk.

[7] The reasons for doing this are twofold. First, females could naturalize through marriage, and not file papers so the female sample is a somewhat selected group likely to be systematically different from the population of all female migrants. Second, after the Act of 1922, females married to a U.S. citizen were not required to file first papers, hence we cannot observe them over time.

[8] Only 15 individuals are observed outside the interval 2-7 years since declaration. Note that immigrants appears in our sample only if their naturalization procedure has been completed - but independently of the citizenship being granted or not. For about 2.6% of the sample citizenship was refused. We keep these records in the sample.

[9] The only possibility is that we cannot find a whole set of documents for a particular petition number. This has occurred only in the 0.2 per cent of our cases.

to derive our key variables, which include occupational scores, popular American names and geographical characteristics.

Defining an American norm and Americanization. We measure conformability with the American norm by exploring how immigrants' names compare with American ones. The first step is to rank American names by the frequency of their appearance in the U.S.-born population which are obtained from the 5% Integrated Public Use Microdata Serie of the 1930 Census (IPUMS, Steven Ruggles et al., 2010). We focus on American individuals born before 1880 and living in the state of New York at the time of the census. The reason behind restricting our attention to individuals aged 50+ is to capture naming patterns for a period that encompasses early migration waves to the U.S. only, hence abstracting from 1930s fashions and naming patterns that might be influenced by the recent migration waves of that time. We also focus on naming patterns in the state of New York only so as to narrow down the pool of names that migrants in the sample were most likely exposed to. Panel B of Table **??** showed common American names. Names including William, John, Charles and George were most common in the 50+ American-born population, which we will consider to be representative of the American norm.

Based on our definition of American norm, we derive a simple metric to capture name Americanization. At any time t we define Americanization as the normalized frequency of names in the 50+ American-born population living in New York state:

$$A_{it} = \frac{S_{itk}}{max(S_{it1}, ..., S_{itK})} \tag{1}$$

$$\text{where} \quad S_{itk} = \sum_k I(Name_{it} = Name_k) \qquad \text{for } k \in 1, ..., K \text{ US-born 50+, NYC} \tag{2}$$

where A_{it} indicates name Americanization, $I(Name_{it} = Name_k)$ is an indicator counting the number of native-born individuals 50+ holding name k and the denominator represents the maximum frequency across all names held by U.S.-born individuals. This bounds name Americanization to be between zero and one.

In other words, A_{it} measures how frequent individual i's name is in the American-born population - defined as before - on a scale from 0 to 1. In fact, names that are unique to our migrants are observed in the American-born population with a frequency equal to zero, so their Americanization will also be zero. On the other hand, migrants who are called William and George will have an Americanization of 1 and 0.68 (given by 0.0612/0.0899, see Table ??), respectively. What is name Americanization? This occurs whenever a migrant changes his name into a name more frequently occurring in the U.S. population. As an example of name Americanization, a migrant called Giovanni who changes his name into William would be Americanizing his name, with a value of A_{i0} equal to zero and a value of A_{i1} equal to one. On the other hand, A_{it} would take a value of one at any t for a British migrant called William who does not change his name. Similarly, if Giovanni were to change his name into Salvatore, A_{it} will take a value of zero for any t, given that neither of these names were common American names. Finally, while Americanizing one's name corresponds to an increase in the index, if migrants were to change into more distinctive foreign-born names, A_{it} will decrease.

This simple metric allows us to define "name changers" as individuals who choose different level of Americanization. The purpose of our index is to capture the distribution of names that is not affected by migration. It differs from other name indices, such as Fryer and Levitt (2004) in one fundamental aspect: while Fryer and Levitt (2004) are interested in a relative index that is invariant to name popularity across minority groups, we aim to measure the popularity of American names only.[10]. Moreover, as shown in Section 6, the overall findings are robust to re-defining popularity using different times and including migrants in the sample of names. This is unsurprising given that the distribution of names has not been substantially affected by immigration and naming patterns were rather stable during the late 1800s and early 1900s.

Measuring occupations and earnings. The typical challenge in studying historical labor market outcomes is the lack of earning measures prior to the 1940 Census. We rely on an indirect, well-established measure of "earning potential" that assigns income scores to each occupation. To obtain the occupational score, we first collected the occupation string from the naturalization papers and standardized occupation titles to match those identified in IPUMS. During this process, all occupations were standardized; however, we flagged those occupations for which some imputation was made for assignment of an occupation (about one-third). Next, we attribute the "occupational score" provided by IPUMS to migrants in our data. Such an income score indicates the median total income (in hundreds of 1950 dollars) of a person in a given occupation. It should be noted that although the occupational score has a well-established use in economics, its limitations are also well recognized (Ferrie,

[10]In our context, the index of Fryer and Levitt (BNI) would be defined as $\frac{\sum_k I(Name_i=Name_k)}{\sum_k I(Name_i=Name_k)+\sum_l I(Name_i=Name_l)}$ (for k in US-born and l in foreign-born), which varies between 0 and 1 as our A_i. As it can be seen, the BNI index weights the frequencies of names by the overall distribution of names across groups. Applying the BNI index to our settings, we find very similar results.

1996, Abramitzky et al., 2012a). For instance, we are unable to measure within occupation-changes in earnings related to the Americanization of a migrant's name. For an effect to be found, it is necessary that changing one's name pushes an individual into a different mean-wage occupation, which annihilates the actual variation in wages. It is possible that our results would be stronger if individual-level information on earnings were available.

Local Labor Markets. An additional and somewhat unique piece of information contained in the naturalization records is the address of residence at declaration and at petition of the migrant and his dependents. We exploit this information in several ways: first, we use it to identify the migrants' local labor market; second, we use it to detect mobility of internal migrants, i.e. those individuals who at the time of petition moved to a local labor market different than at declaration; and third, at an even more disaggregated level, we control for the potential role of ethnic enclaves.

We implement a geocoding procedure through which we use migrants' addresses to derive coordinates, subsequently assigning local labor markets to each individual. Local labor markets were defined as those geographic areas which provided a better fit across several regressions where individual labor market outcomes were modeled as a function of individual level characteristics and several alternative definitions of geographic entities. Our preferred choice for labor market falls with the community districts.[11]

Summary Statistics Figure 1 in the introduction already showed the positive association between name Americanization and occupational scores. Table 2 further explores the dimensions along which migrants who Americanize their names and those who do not differ, both at the time of declaration and between declaration and petition.

INSERT TABLE 2 HERE

As shown in the first column, at declaration migrants are about thirty years old, have been in the U.S. for about seven years and about 45% of them are married (although only 32% of them have been joined by their spouse in the U.S.). About 30% of the migrants have children. A large fraction were born in Italy, Russia or other major sending countries of this time, such as Poland or the Czech Republic. We call these migrants "new migrants" in

[11] Exploiting the detailed geographical information, it is easy to note that Brooklyn and Manhattan were the two most populated boroughs. Distinguishing the place of residence by major country of birth, some clusters by country of origin are evident, including those of Italians in the lower east side of Manhattan and Russians in the Brownsville area of Brooklyn, yet different ethnic groups did coexist in the same areas. Detailed explanations of the construction of these local labor markets, as well as results and additional tables are reported in the online Appendix.

contrast with the group originating from Northern Europe, who arrived in the U.S. during the first wave of Mass Migration, at the beginning of the 1800s. The classification between old and new migrants distinguishes ethnic groups who have an established history of successful settlement in the U.S. from new waves of migrants that ignited xenophobic reactions in the early 1900s (Hatton and Williamson, 2004). As we expected (given the time period of analysis), approximately 60% of the naturalized in 1930 came from new sending regions.

The other columns of Table 2 show the statistics by name Americanization. Name Americanization was quite common: there are 1,017 individuals in the sample who Americanized their name. Both groups of migrants who changed and who kept their names exhibit essentially the same average characteristics in terms of age, marital status, probability of having children and occupational score measured at declaration time. Differences arise primarily in three traits: first, at the time of petition, migrants who Americanized had stayed longer in the U.S. compared to name keepers; second, they had a higher probability of having their spouse present in the country, and third the distribution by country of birth varied substantially. New migrants are far more likely to Americanize their names than old migrants. Part of this is likely driven by their names being less common in the U.S. than names of old migrants from the UK or Ireland. This can be seen, for instance, in the fact that (from among the new migrants) Russians are far more likely to Americanize their names than Irish. All these patterns persist when looking at quartiles of the Americanization index.

The second panel of Table 2 shows changes in the characteristics of our sample over time. There are very few differences across groups. Migrants who Americanize and those who do not Americanize their names do not differ in terms of the timing of petition filing, and therefore, despite the original gap in years since migration, both groups file a petition within five years of having filed first papers. This behavior is also reflected in the change in age. Importantly however, the two groups show substantially different raw means in terms of labor market outcomes: changes in the occupational score occurring between declaration and petition are essentially zero for name keepers and for those who changed into a more distinctive name, but are positive for migrants who Americanized (0.064 log points on average). Such payoff is observed across nearly all quartiles of the Americanization index.

3 Empirical Model

The summary statistics highlight the correlation between name Americanization and occupational standing. Yet, part of this payoff might be driven by observable differences between individuals with different level of Americanization. More importantly, those who Americanize and those who do not might differ across other - perhaps unobservable - dimensions.

Going beyond a simple descriptive analysis, we present below a number of empirical strategies used to understand whether this correlation persists once compositional differences are taken into account.

The OLS estimator. Our first model considers a regression equation relating occupational scores and name Americanization:

$$
\begin{aligned}
y_{it} = {} & \beta_0 + \beta_1 A_{it} + \beta_2 Trend_{it} + \boldsymbol{x'_{it}}\gamma + \\
& + \beta_3 \Big(Trend_{it} * COB_i\Big) + \beta_4 \Big(Trend_{it} * LabMkt_i\Big) + \tau_i + \epsilon_{it}.
\end{aligned}
\tag{3}
$$

where y_{it} is the log-occupational score of individual i at time t. A_{it} is our key variable, as defined in equation (1), and representing the normalized frequency of individual's i name in the U.S. born population at time t.

We also include a time variable, $Trend_{it}$. As not all migrants file first papers at the same time, a linear time trend is individual specific and captures the different time periods of observation in the data. This trend has both economic and econometric relevance, given that it controls for different immigrant cohorts and changes in the occupational distribution that are also driven by the business cycle.

$\boldsymbol{x'_{it}}$ is a vector of time varying socio-economic variables, such as marital status or having children.

Equation (3) also shows two interaction terms: one between the time trend and indicators for the country of birth ($Trend_t * COB_i$) and one between the time trend and indicators for the local labor market ($Trend_t * LabMkt_i$). Controlling for country of origin and time trends in occupational scores seems important in our context. For instance, Abramitzky et al. (2012a) reports considerable variation in occupational score patterns across nationalities over time. Furthermore, several documents of this time report episodes of exclusion of aliens from engaging in particular occupations.[12] Hence, occupational changes might have followed different trajectories depending on the individual's country of birth. Similarly, the interaction terms $Trend_t * LabMkt_i$ controls for different labor market patterns that might simultaneously affect occupation score and name changes.

Finally, the term τ_i is a vector of time-invariant and unobservable personal traits that have an impact on occupational scores (such as innate ability) and may be correlated with other regressors in the model. ϵ_{it} is an error term satisfying the usual assumptions.

This first estimation strategy simply pools the data and applies OLS estimation to equa-

[12] For example, in 1914, the Supreme Court upheld a New York statute excluding non-naturalized Italians from engaging in building the New York subway; and in the 1920s, licensing acts in many states barred aliens from practicing highly qualified professions (Higham, 2002).

tion (3). By doing so, we control for variables such as personal and labor market attributes to examine the existence and extent of payoffs associated with acquiring American names. If name Americanization were fully explained by observable differences across individuals, controlling for them would eliminate the potential bias in our estimates which stems from individual self-selection with the OLS results pinning down the extent to which name Americanization and occupational scores are related.

A first-difference estimator. There might be several unobservable traits such as individual time-invariant ability (represented by τ_i in equation labelbasemod) and correlated with name Americanization and occupational scores that might drive our OLS results. We therefore proceed by adopting a second estimation method.

Re-writing equation (3) for occupational scores observed in the declaration documents as:

$$\begin{aligned} y_{iDecl} = \beta_0 + \beta_1 A_{iDecl} + \beta_2 Trend_{iDecl} + \boldsymbol{x}'_{iDecl}\gamma + \\ + \beta_3\Big(Trend_{iDecl} * COB_i\Big) + \beta_4\Big(Trend_{iDecl} * LabMkt_i\Big) + \tau_i + \epsilon_{iDecl}. \end{aligned} \tag{4}$$

and for occupational scores observed in the petition documents as:

$$\begin{aligned} y_{iPet} = \beta_0 + \beta_1 A_{iPet} + \beta_2 Trend_{iPet} + \boldsymbol{x}'_{iPet}\gamma + \\ + \beta_3\Big(Trend_{iPet} * COB_i\Big) + \beta_4\Big(Trend_{iPet} * LabMkt_i\Big) + \tau_i + \epsilon_{iPet}. \end{aligned} \tag{5}$$

Taking the difference between these two equations yields a model in first differences:

$$\begin{aligned} (y_{iPet} - y_{iDecl}) = \beta_1(A_{iPet} - A_{iDecl}) + \\ \beta_2(Trend_{iPet} - Trend_{iDecl}) + (\boldsymbol{x}_{iPet} - \boldsymbol{x}_{iDecl})'\gamma + \\ + \beta_3\Big((Trend_{iPet} - Trend_{iDecl}) * COB_i\Big) + \\ + \beta_4\Big((Trend_{iPet} - Trend_{iDecl}) * LabMkt_i\Big) + (\epsilon_{iPet} - \epsilon_{iDecl}). \end{aligned} \tag{6}$$

Here, $(A_{i,Pet} - A_{i,Decl})$ is a continuous variable bounded between -1 and 1, indicating how much more or less frequent the migrant's name becomes over time in comparison to the U.S. population. From our previous examples, this variable would take a value of one for Giovanni who Americanized his name to William, and zero for both the British William who did not change name and the Italian Giovanni who changed name into Salvatore.

Due to the structure of our data we need a modification of (6). Equation (6) correctly measures the association between changing names and occupational scores if all migrants change name between declaration and petition. If migrants were to Americanize their names before declaration, the previous equation would consider these migrants as name keepers. For instance, if Giovanni were to become William after arrival yet before declaration, $(A_{i,Pet} - A_{i,Decl})$ would take value of zero as we would observe Giovanni holding a common American name in both time periods. However, if an association exists between names and occupational scores, y_{iDecl} would have already been influenced by the new name, i.e. Giovanni's y_{iDecl} would have changed in response to the new name. In order not to confound name changers and name keepers, we substitute A_{iDecl} with $A_{iArrival}$, which is based on the name reported in the certificate of arrival. Through this implementation, the "will-be-William" Giovanni is considered as having Americanized his name, as opposed to being a name keeper. Using name at arrival is arguably more exogenous to unobserved U.S. shocks that might occur after arrival and potentially provides a more conservative estimate of the relationship of interest. In estimating equation 7 we are implicitly assuming that there is no dynamic selection into name Americanization. In other words, we assume that individuals who change name between arrival and declaration are observationally similar to those changing names between declaration and petition after controlling for relevant characteristics. This assumption is used by Arai and Thoursie (2009) and allows us to substitute A_{iDecl} with $A_{iArrival}$. [13]

To summarize, our second estimation method is a modified first-difference technique (FD) based on the following model:

$$\begin{aligned}(y_{iPet} - y_{iDecl}) = {} & \beta_1(A_{iPet} - A_{iArrival}) + \\ & \beta_2(Trend_{iPet} - Trend_{iDecl}) + (\boldsymbol{x}_{iPet} - \boldsymbol{x}_{iDecl})'\gamma + \\ & + \beta_3\Big((Trend_{iPet} - Trend_{iDecl}) * COB_i\Big) + \\ & + \beta_4\Big((Trend_{iPet} - Trend_{iDecl}) * LabMkt_i\Big) + (\epsilon_{iPet} - \epsilon_{iDecl}).\end{aligned} \tag{7}$$

In order to keep our notation simple, we use ΔA_i to indicate such difference in the remainder of the paper.

Compared to a simple pooled OLS estimator of (3), equation 7 offers the advantage that we are able to purge out of the model time-invariant and unobservable personal traits (such as innate ability) that might have an impact on occupational scores and the propensity of name Americanization, while still controlling for specific trends in earning growth by country of birth and local labor markets, aimed at reducing potential biases due to occupational or geographical sorting.

If these characteristics were sufficient to guarantee conditional independence of name

[13] However, we perform several robustness checks with respect to this definition in Section 6 and further details are available in the online Appendix.

Americanization from the unobserved component of the model, we would conclude that these results estimate a causal effect of name changes on occupational scores. However, it is possible that several relevant factors might still be omitted in the regression analysis, that reverse causality might be at work and that time varying self-selection might be present. For example, migrants who Americanized might have accumulated different levels of human capital or language ability. Additionally, name Americanization might occur before occupational change for some individuals while occupational change might lead to name change for others. The last two estimators we adopt aim to control for time-varying factors that might be correlated with name Americanization and occupational upgrading.

Focusing on name-changers only. With a third method, we restrict our sample to name changers only (NC), exploiting the different timing of name choices. By using this strategy, the control group for name changers consists of those who did not change name at the time of declaration. This should better control for time-varying individual factors that might cause selection (a similar strategy is followed in Arai and Thoursie, 2009). For instance, as long as name changers exhibit similar patterns of human capital accumulation or language acquisition, this strategy will net out the effects of these variables.

An IV estimator. While focusing on name changers only could reduce endogeneity concerns, identification still relies on the assumption of parallel trends between current and future name changers, after conditioning on the full set of characteristics. Furthermore, this method does not rule out the possibility of reverse causality. To recover a causal link between name Americanization and economic success, we apply an instrumental variable (IV) strategy on equation (7).

In order to understand the causal link between changing name and economic success, we need a variable that predicts the former without directly influencing the latter. We look for a measure of linguistic complexity that is unrelated to the specific content of a name - as the latter might be correlated with socioeconomic outcomes and, hence, ultimately correlated with occupational scores.

We predict name Americanization as a function of the simple word points the name at entry would receive from the popular board game Scrabble. The origin of the Scrabble point system dates to 1938, and the point values were worked out by the architect Alfred Moscher Butts by performing a frequency analysis of letters from the front page of various newspapers. A name associated with relatively high points would correspond to a name of an unusual linguistic complexity or euphoniously uncommon, whereas a name with low points would be constituted of relatively simple phoneme.

In practice, we create a measure of distance between the scrabble points of the migrant's name at arrival and the scrabble points associated with the "American norm". Our Scrabble index $S_{jArrival}$ is defined as:

$$S_{jArrival} = \frac{SP_{jArrival}}{\sum_{w \neq j} SP_w / \sum_{w \neq j} 1}, \tag{8}$$

where $SP_{jArrival}$ is the scrabble points of name j, divided by the median scrabble points across all American individuals born before 1880 and currently living in the state of New York (Figure 3(a)), excluding name j from the computation of the median. Values of S_j substantially above or below 1 indicate a name that is "phonetically" distant from the American norm. This normalization allows us to measure the degree of complexity of the arrival name with respect to the available linguistic complexity at destination.

INSERT FIGURE 3 HERE

The Scrabble index appears a suitable instrument for several reasons. Given that points are based on the frequency of letters, the Scrabble index does not make use of the semantics associated with names, nor their length. For this reason, it is unlikely that this index is associated with individuals' labor market performance and thus our exclusion restriction should be valid. While we cannot directly test for the validity of the exclusion, Table **??** in the online Appendix reports summary statistics of our sample by separating our group into those who have below and above median values of S_j. The table suggests that both groups of individuals are observationally similar. In particular, there are no differences in terms of years since migration, which means that the instrument is capable of purging any channel linked to human capital accumulation, including acquisition of language skills. Furthermore, Aura and Hess (2010) show that Scrabble points are generally unrelated with lifetime outcomes, and, particularly, do not predict occupational prestige during 1994 and 2002.

The remaining panels of Figure 3 show how our Scrabble index predicts name Americanization. Figure 3(b) shows the graphical distribution of $S_{jArrival}$; the vertical line indicates the median, which is very close to one. In the same graph, we report selected values of ΔA_i associated with the Scrabble index (for ease of representation, we calculated means of Scrabble index and Americanization index by aggregating for each name at petition). The blue markers refer to ΔA_i which are above the mean; the remaining are values of ΔA_i pertaining to observations from the lowest decile. The first noteworthy aspect is that the distribution of S_j at both points in time is bell shaped, indicating that a larger mass of people have names of average complexity in the U.S. Second, and more importantly, there is a clear relationship between the distribution of $S_{jArrival}$ and ΔA_i. Highest values of Americanization are found within an interval of the Scrabble index corresponding to the American norm. On the other hand, lowest values of Americanization are scattered further away, towards both tails of the distribution. This means that the Scrabble index could be considered as a functional predictor of Americanization. Another way of interpreting this is to observe the "dynamics" of the Scrabble index for individuals who Americanize. In Figure 3(c), we show the distribution obtained by subtracting $SP_{jArrival}$ from the Scrabble points calculated at the time of petition, when migrants changed name. We omit from the graph name keepers (i.e. those for which $SP_{jArrival}$ did not vary over time). The distribution shows that the vast majority of changes in the Scrabble index happen in an interval around 0.5 points from the American norm. This means that only those close to the "Scrabble norm" find it beneficial to change name. For those who are too far from the norm, the "costs" of Americanizing their name are so high that we observe very few instances of name changes, and a negligible increase in the Americanization index. Remarkably, the same pattern emerges when we look at this "difference" distribution by country, as well as by aggregating by popular names of migrants at petition (see Figures **??** and **??** in the online Appendix for additional details).

The graphical representation above suggests the existence of a non-linear relationship between S_j and ΔA_i. We now test the validity of the instrument in the second panel of Table 4, while the first panel shows the second stage estimates.

While the IV estimator has the advantage of showing whether the positive association found in the summary statistics persists after controlling for time-varying, individual-specific factors, it is important to remind that estimates should be interpreted as measuring a local average treatment effect. This is the average effect of changing names on occupational scores for complier migrants, i.e. those who change their name because of particular values of the instrument, but who would not have changed their name otherwise. The interpretation of our estimates does not extend to individuals who (do not) change name irrespectively of the instrument, i.e. "always-takers" ("never-takers"). While the LATE estimates might be different from the average treatment effect, this strategy will still be informative on the existence of a causal relationship between name Americanization and occupation scores and provides additional evidence on the direction of biases.

The OLS estimator, the modified first-difference estimator (FD), focusing on name changers only (NC) and the instrumental variable method (IV) will show whether name Americanization had a return in the labor market, i.e. what were the consequences for such choices.

4 The Effect of Changing into a Popular American Name

We start the analysis in Table 3, where the first three columns show the estimates in levels (i.e. pooling equations 4 and 5), ignoring individual heterogeneity, while the last three columns show the estimates of equation (7). Across columns we vary the set of regressors given in our empirical framework. These range from migrant and household characteristics (whether or not married, whether or not children present and whether or not spouse present) to country of birth and geography (community districts at declaration).

INSERT TABLE 3 HERE

The OLS estimator. In the pooled OLS regression, name Americanization is associated with a 4% increase in occupational score, a payoff that remains stable across the three specifications. Having a spouse present in the U.S. is associated with higher occupational scores, while marital status or having children does not seem to have an impact on the migrant's economic outcomes. Note that these variables aim at capturing time-varying characteristics that might affect assimilation incentives - and hence occupational upgrading. In column (II), we condition on nationality-specific indicators to capture different labor market patterns across origin groups. We find that our results become statistically weaker and point estimates slightly decrease. Nonetheless, the association between Americanization and occupational upgrading remains positive even within nationality-specific occupational scores patterns. Compared to the group of old migrants (which includes British, Swedish and Norwegian migrants), Germans exhibit higher labor market success, while the remaining groups are under-performing.

Lastly, we explore the influence of local labor markets by including in Column (III) indicators for community districts where the immigrants live at the time of declaration. The pattern of our baseline results is unchanged and we find a statistically significant return to changing own name to common American names. Despite some evidence of sorting from Figure **??**, this seems to indicate that the impact of local labor market conditions is relatively mild and does not drive our main findings.

The first-difference estimator. Exploiting the panel structure of the data, the last three columns of the table report the estimates in first differences. The positive association between changing into a popular American name and changes in occupational scores remains, with labor market payoffs being larger the more popular the name, and becoming larger in magnitude and more statistically significant. Conditioning on a number of variables, we find that changing from a non-American name to a very common American name is associated with a 14% return. Household characteristics have little impact on occupational upgrading. Nationality specific trends indicate the presence of little occupational convergence, except for the Irish and the Other migrant group.

Assuming that individual self-selection is purely driven by observable traits we control for, or by time-invariant characteristics, these estimate indicate that the positive association found in our OLS analysis is not driven by selection but is likely driven by the market rewarding individuals with an American name. In fact, name Americanization results in a 14% increase in occupational scores when individual productivity is unchanged.

INSERT FIGURE 2 HERE

Figure 2 further summarizes this finding, namely that there is a positive association between name Americanization and occupational upgrading, with payoffs being larger the more American the name. For example, changing a name into William or John is associated with a 14% increase compared to keeping one's name. On the other hand, adopting a not-so-common American name such as Hyman or Isidore has a much lower - statistically zero - payoff.

INSERT TABLE 4 HERE

Focusing on name-changers only. Within the first-difference estimator, we restrict our sample to name changers only, i.e. exploiting the different timing of the name choice. We estimate the model progressively adding regressors as in our baseline specification. Results are shown in the first three column of Table 4. While sample sizes are smaller, patterns are quite remarkable: we still find a positive relationship between acquiring popular American names and economic success in the U.S. As long as name changers in one year have similar trends to those in another year, these estimates capture the causal effect of name changes on occupational scores. Furthermore, as long as time-varying unobservable characteristics are shared within this group, these estimates implicitly control for variables such as English acquisition and human capital accumulation which are unobserved to us.

The IV estimator. In the first stage, we start by using our Scrabble index as a simple linear predictor of Americanization in columns IV to VI. In columns VII to IX, we model the non-linear relationship we had seen descriptively in Figure 3, adding a quadratic term. Throughout our models, $S_{jArrival}$ is positively associated with name Americanization. The quadratic term is also highly significant, hinting at a concave relationship between the scrabble index and name Americanization, even after controlling for all the covariates. This suggests that individuals who Americanize have higher scrabble points, although incentives to Americanization decreases as names become further away from the norm. The instruments continue to perform well independently of the controls added, and the Scrabble index remains a valid predictor, as shown by the first stage F-statistics which indicate the validity of our instrument. The quadratic model in Scrabble index allows us to test the validity of our exclusion restriction via a Sargan test for overidentifying restrictions. We cannot reject the null of the instruments being valid exclusions. Finally, we tested whether there is evidence of endogeneity. While we cannot fully reject exogeneity of name Americanization in the first model, in the second we reject such a hypothesis at the five per cent significance level.

After testing the validity of the instrument, we find in the second stage that the returns to name Americanization increase once we instrument for A_{it}. It appears that controlling for the possible time-varying self-selection of individuals as well as any other form of measurement error into the decision to change names increases the return to name Americanization to about 70%. As mentioned earlier, this should be interpreted as the average effect of name Americanization on occupational scores for complier migrants, i.e. for those who change their name because they have a rather high Scrabble index, but who would not have changed their name had they had a low Scrabble index. Being a local average treatment effect, it is unsurprising that the point estimates are relatively higher. While potentially far from the average treatment effect, these estimates further corroborate the existence of a causal relationship between our variables of interest.

In summary, the positive relationship between name Americanization and occupational upgrading persists after tackling directly the endogeneity of Americanization with our IV approach. Our IV results suggest that it is the act of Americanization per se that causes higher occupational upgrading, and not (only) the selection of migrants into such choice, since the instrument allows to "changing" at random the incentives to Americanize.

5 Channels

The existence of positive returns to name Americanization poses two important questions. The first one is why not everybody Americanized their names. The second relates to the the expected direction of self-selection.

There might be various reasons why only some migrants Americanized their names. Fryer and Levitt (2004) list several frameworks that help explaining persistence of distinctive names. First, individuals might not have full information on the returns to name Americanization. Second, within their price theory model of names, incentives to name Americanization will differ by neighborhood composition. For example, if moving costs across neighborhood fall, name Americanization should occur more frequently. Similarly, within ethnic enclaves the benefits from name Americanization might be lower and the costs could be higher. Third, individuals might have utility gains if they follow a prescribed behavior of a particular group (Akerlof and Kranton, 2000). Hence, only those individuals subject to few identity prescriptions will Americanize their names.

While these frameworks suggest that migrants who Americanize their names should be observationally different than those who do not, it is a priory unclear whether Americanization occurs among migrants facing occupational mobility or among those who do not. For instance, in presence of imperfect information, higher ability migrants would be probably more likely to Americanize their names. In the context of a price theory model, individuals in more segregated neighborhoods might face a higher cost of Americanization, if there is a cost for holding an American name when interacting with migrants. Yet, the type of individual who Americanize his name will depend on whether it is the high or low skilled workers who live in ethnic enclaves. Lastly, within the identity framework, it is a priori uncertain whether behavioral prescriptions would be stronger among those who are successful in the labor market. Overall, the type of migrant who Americanize cannot be pinned down univocally.

Our results highlight that individual self-selection did play a role in name Americanization beyond the simple observable differences across individuals. A comparison of the OLS estimate (4%) with the FD (13%), NC (13%) and IV (70%) estimates highlights an interesting and consistent direction of the bias of the OLS estimator: there is negative selection on time-invariant (FD) and time-varying (NC, IV) unobservable characteristics. This suggests that the OLS returns to name Americanization might be a lower bound of the true payoff from this choice. Hence, it was the migrants facing the greatest barriers to occupational mobility who Americanized their names.

This result hints to the fact that imperfect information might not be the driving factor of selection into name Americanization. This evidence is also consistent with Goldin and Shim (2004), who also finds that keeping the own surname is more common among the highly skilled women. Likely high achieving individuals did not need to change name as their quality was already signaled in the market, as they were subject to less discrimination, or as they could find alternative ways for occupational improvements such as migration or human capital accumulation.

We further corroborate the hypothesis that name Americanization was a way to overcome labor market barriers by examining the sub-populations in which it had a higher return. We would expect higher payoffs for groups that were lower skilled, potentially more discriminated against, or with less alternative means for socio-economic improvement.

INSERT TABLE 5 HERE

Table 5 breaks the sample into several categories. We report estimates of the payoff to name Americanization across all estimators for new and old migrants, for internal migrants and internal stayers and for individuals who live in ethnic enclaves.[14]

In the first panel we distinguish the effects for migrants whose first settlement happened during the early wave of mass migration (Germans, Scandinavians, Irish, British) and for those who came from regions with a later experience of emigration to the U.S. (Italians, Eastern Europeans, Russians). The pattern of results for the new migrants is similar to our baseline estimates. For the old migrants, the magnitude of estimates is smaller and the effect is statistically insignificant throughout. It appears that changing into a popular American name determined a premium for the "new" migrants only. These heterogenous returns could be driven by stronger discrimination in the labor market for this group, or by their relatively low level of skills. Historical accounts (e.g., Higham, 2002) seem to confirm that there were widespread resentment and discrimination against the new migrants.

In the second panel, we perform our analysis by focusing on individuals who changed labor market between declaration and petition (internal migrants) and those who do not. Internal migrants exhibit a much smaller and statistically insignificant return to name Americanization, while the payoff is particularly high for those who did not move. These results are compatible with an interpretation of migration as a form of human capital investment. Acquiring skills through migration can be seen as a strategy of earning growth that is competing with changing name. Therefore, returns to Americanization will be smaller for internal migrants. Furthermore, if internal migrants are positively selected - as arguably only those who can afford it will migrate - the higher payoff for non-migrants is again indicative of name Americanization as a way to circumvent lock-in effects for those who had no alternatives means for occupational mobility.

The last panel of the table explores the role of ethnic enclaves. We calculate a measure of exposure to ethnic enclaves using the 1% sample from the 1920 IPUMS. Exposure is defined as $\frac{P_{kh}}{P_h}$ where P_{kh} represents the number of individuals living in the census tract h who were born in country k and P_h is the total number of individuals – including U.S.-born – in each tract. This index provides a measure of the relative size of each immigrant group in the locality and has been previously used to study the role of enclaves (e.g. Borjas, 2000). We focus on a measure pertaining to the year 1920 in order to capture an ethnic composition

[14] In the online Appendix, we perform additional analysis by physical characteristics and occupations. All results are consistent with our interpretation.

that is pre-determined to the decision of changing name. We then divided migrants living in neighborhoods with high and low ethnic concentration. We create an indicator for the share of own nationals being in the top quartile of the distributions. For instance, we are comparing Italians living in little Italy with those living in Brooklyn.[15]

The estimates suggest that name Americanization did not pay for individuals living in enclaves. On the other hand, migrants in less segregated neighborhoods had a higher return to name Americanization. This indicates, once again, that name Americanization paid for those individuals who were somehow more constrained and less likely to interact with co-nationals.

6 Robustness Checks

Before concluding, in Table 6, we carry out a series of checks aimed at ascertaining the robustness of our results to various data aspects and definitions.

INSERT TABLE 6 HERE

Our model in equation (7) uses a definition of name change based on the difference between name at petition and name at arrival. While the petition was the document that officiated the new name, such a change could have occurred at different periods prior to petition which we can to some extent identify. To be more precise, Figure 4 shows the different timings of possible name changes in relation to the timing of documents. Recall that we collect information from all naturalization papers, including the certificate of arrival (which has limited information apart from the arrival name of the person) and the declaration of intention, as well as the petition for naturalization.

INSERT FIGURE 4 HERE

Each line in the figure represents an individual "type". Dotted lines indicate periods when the name popularity corresponds to the first name observed when entering the U.S., as written on the certificate of arrival ($Name_{iArrival}$). Solid lines represent a change in name compared to the original name. There were no cases in the dataset where the name is changed twice into two different names. However, there are several "types" of individuals. For instance Type 1 individuals are those who never change their names (around 54.13%

[15]Since we are only able to match the census tracts for Manhattan and Brooklyn at the time of writing, the sample size in this exercise is limited to immigrants living in these two boroughs, which means that those living in Bronx, Queens, Staten Island and outside New York City (internal movers) at the time of declaration are excluded.

of the sample). Type 2 individuals change name between arrival and declaration (around 75.81% of the name changers), while Type 3 individuals change name between declaration and petition (around 24.19% of the name changers). Lastly, we also observe a fourth type of individual who is observed to have changed name between arrival and declaration yet reports the original name at the time of petition (around 3.62% of the name changers). It is unclear from the data whether these individuals had been temporarily using the "new" name or whether the new name had been 'imputed' or 'misunderstood' by the clerk. Therefore we checked all these individuals' name and compared the reported name with the name in the signature. Type 4 individuals who signed the declaration with their original name were considered as name keepers while the rest are probably individuals who really changed their name multiple times.[16]

The data challenge derives from the fact that, although we observe names at three points in time, occupations are only observed at declaration and petition. In an ideal dataset, we would observe individual names at the first occupation after arrival (prior name change), comparing these with name at arrival, and observing these individuals over time. However, our data do not contain information on the first occupation upon arrival and prior-name change. To our knowledge, no other source can be used to detect this information, as passengers records report either occupation at origin or potential occupations in the U.S. (Ferrie, 1999)[17]

Our baseline model considered changes in name between petition and the time of arrival, $(A_{iPet} - A_{iArrival})$, to capture all the name changers in the sample. By this definition, Type 4 individuals are considered non-changers. As mentioned in previous sections, this definition offers the advantage of defining a name change based on a variable that is pre-determined with respect to any possible post-arrival outcome in the U.S. Therefore, this definition reduces potential concerns of reverse causality.

While the benchmark model correctly identifies Type 2 individuals as being 'treated' by the change of name, their occupational scores have already been affected by such change at the moment of declaration. If name Americanization provided a one-time increase in occupational scores, the inclusion of these individuals in the treatment group would bias downward the baseline results, as occupational scores' trajectories would have already changed for them. Our baseline definition would hence provide a lower-bound effect. On the other hand, if name Americanization implied a change in both average occupational scores and growth,

[16]Results remain unchanged if we drop Type 4 individuals.

[17]If a larger sample size was available it would be possible to match the migrants of our dataset with the Census enumeration and obtain a third point in time. We would need, however, enough observations of individuals who arrive prior 1920, do not change name by the 1920 Census, and declare between 1920-1930. For the time being, this route did not seem feasible.

with the inclusion of Type 2 individuals we would still be able to detect a return to name Americanization, although we would not be able to identify the short-run change from the long-run change in occupational scores. However, under this second scenario, the association between names and occupational scores would still be detected by our analysis. Accordingly, these are the reasons why we have preferred to keep the definition based on name at arrival in our baseline results, as it is potentially rather conservative and provides a larger sample size.[18]

We adopt two strategies to examine whether our preferred empirical model delivers conservative estimates. First, we drop the Type 2 individuals and therefore estimate a true first-difference model.[19] Second, we restrict our sample to individuals who declare their intention to naturalize within three years from arrival and drop everybody else in the sample. By doing so, although we cannot fully control for the fact that Type 2 migrants have already been 'treated' compared to Type 3 migrants, we can at least limit the effects of new names to a shorter time span. Furthermore, we are able to better capture the trajectories immediately after arrival and, presumably, work with a more exogenous initial condition for the occupational distribution.[20]

Results in Table 6 are consistent with previous estimates. Point estimates are higher if only changers between declaration and petition are kept. It should be noted that given the small number of type 3 individuals, it is probably unsurprising that the estimates for NC and IV are very imprecise. On the other hand, restricting the sample to those early declarants does not particularly alter our main conclusions, in that we still find a positive relationship between changing name and occupational upgrading.

Lastly, we run several analyses to ascertain the sensitivity of our results to the definition of Americanization. When looking at historical records, it is well known that many orthographic differences might appear in first and last names. In such circumstances, it might be important to standardize names using phonetically equivalent transliterations. Common algorithms to obtain name standardizations are Soundex, NYSIIS, Metaphone, and Caverphone. In the context of our study, it could be the case that a misspelling between our data and the Census causes the Americanization level to be measured with error. For example, if the name John is misspelled as Jon in our sample, the migrant would be considered to have Americanized his name when in fact the name had not changed. We report our baseline results by considering

[18] Further details in the online Appendix.

[19] We wish to stress, once again, that we are aware that this will not induce additional bias only under the assumption that, conditional on characteristics, the timing of name change is exogenous, or otherwise put, that name changers at different point in time have parallel trends.

[20] We want to stress once again that reporting a first-difference model on the whole sample would not be meaningful as we would then consider as "untreated" the migrants who changed their names between arrival and declaration.

a name change to be a change into a phonetically different name, thus reducing differences deriving from simple misspellings, yet also requiring name Americanization to be a change in the "sound" of the name. We use the Soundex algorithm. Names that sound the same yet are spelled differently, such as John and Jon, have the same code. Therefore we are able to purge possible orthographical errors made in the original record by the court clerks as well as during our data collection process. As can be seen, results remain within the confidence interval of our baseline analysis, even when this restrictive definition is adopted.

The final four columns change the definition of American norm, by constructing the A_{it} variable using the name distribution of the full population living in New York in 1930. Changing the norm definition does not alter our conclusions. [21]

7 Conclusions

This paper provided the first direct evidence on the magnitude and consequences of name Americanization. Such phenomenon, known only anecdotically, was not only widespread, with an average of 30% of migrants adopting popular American names, but had also substantial impact on upward mobility during the first half of the century. We find, in fact, that immigrants who Americanized their names experienced higher occupation-based earnings growth than those who did not. These results persists across all our specifications, including those controlling for individual heterogeneity or those based on an instrumental variable technique where we use a Scrabble index that reflects the linguistic complexity of migrant names when they enter the U.S. as an instrument for name Americanization.

These stark differences stemming from name Americanization provide important insights for current work.

Our unique data source is informative for those current studies based on record linkage. Linking individuals by name, age and birthplace has become a common methodology in constructing longitudinal historical data. For example, Abramitzky et al. (2012b) estimates the return to migration by linking Norway-to-US migrants with their brothers who stayed in Norway in the late nineteenth century. Abramitzky et al. (2012a) studies immigrant assimilation by linking migrants and natives in the Censuses over time. Long and Ferrie (2013) links fathers and sons in the 1850 to study occupational mobility in the U.S. and Britain.

[21] In the online Appendix, we run further checks to understand the sensitivity of our analysis to coding errors. For instance, we drop the flagged occupations, we exclude the addresses that we imputed and we further explore the sensitivity of the benchmark results by changing the definition of the local labor markets. None of these checks change our conclusions. The Appendix also reports additional channels of heterogeneous effects, such as surname Americanization, employment types, age or occupations particularly hit by the Depression.

A name change in the U.S. - especially in the studies focusing on immigrants - might be a cause for a failed match. We have shown in our analysis that name Americanization was more prevalent among migrants facing stronger barriers in the labor market (lower earnings trajectories, more likely to be coming from the new sending areas), yet occupational upgrading followed name Americanization. While these opposite features of the Americanization process - negative selection, but subsequent higher earnings growth - do not clearly suggest implications for estimates based on matched sample, they are at least informative in showing that immigrants who Americanized their names[22] are not observationally similar to name keepers on both observable and unobservable traits. Hence, particular care should be put in understanding the representativeness of linked samples.

Furthermore, our results highlight the existence of a tradeoff between maintaining individual identity and enhancing labor market success, suggesting that the assimilation process in the first half of the 1900s passed through economic as well as cultural assimilation. In an historical perspective, this implies that despite migrant occupational upgrading being limited (Abramitzky et al., 2012a), migrants adopted alternative strategies to climb the occupational ladder. Returns to name Americanization were high. While the simple OLS estimates suggest as potential explanations for the association between name Americanization and earnings to be changes in the market valuation of migrants' skill endowment, reduction of discrimination, or more rapid human capital accumulation, the first-difference and IV estimates rule out many of these reasons, with the exception of discrimination. Consequences are twofold. First, the low occupational mobility observed in previous studies might have been caused by different attitudes and discrimination levels towards different ethnicities. Second, in a broader perspective, the implied tradeoff between own identity and labor market success suggested in several recent analysis (e.g,. Bertrand and Mullainathan, 2004, Fryer and Levitt, 2004, Arai and Thoursie, 2009) seems to have been present since the early making of modern America.

The online Appendix can be found at the following link: `https://www.dropbox.com/s/sjh8boldotau8qy/appendix_10212013.pdf`.

[22] Surname could also matter. We have mentioned that surname Americanization occurred only in 7% of our sample. However, surname changes occur in 40% of the sample, while surname changes based on a Soundex transliteration of the surname in 30% of the sample.

8 Tables

Table 1: Popular American Names and Name Americanization

A) Name Americanization		
Country of origin	% Americanized	N
Italy	0.189	729
Russian Empire	0.567	598
Other new migrants	0.467	683
Germany	0.258	365
Ireland	0.038	316
Other old migrants	0.130	385
Other	0.235	277
Total	0.303	3353

B) Popular American Names		
Name	Frequency	Cumulative
William	0.0899	0.0899
John	0.0891	0.1790
Charles	0.0654	0.2444
George	0.0612	0.3056
...	...	...
Otto	0.0014	0.7915
...	...	...
Angelo	0.0001	0.9324

Panel A): own elaborations from sample of naturalizing immigrants in 1930, Ancestry.com. Name Americanization is defined as the change into a name with a higher frequency among popular American names. Panel B): own elaborations from IPUMS Census, 1930. Name frequencies are calculated using US-born population aged 50+ in 1930 New York.

Table 2: Characteristics by level of Americanization: at Declaration and differences between Petition and Declaration

Variable	All	Distinctive	Keepers	Americanize	1^{st}	2^{nd}	3^{rd}	4^{th}
	At declaration							
Log Occupational Score	3.1509	3.2185	3.1446	3.1579	3.1522	3.1825	3.1668	3.1313
	(0.4679)	(0.4351)	(0.4339)	(0.5376)	(0.5345)	(0.5311)	(0.5781)	(0.5073)
Age	29.9039	28.6471	30.4471	28.8396	28.7556	28.3049	27.8560	30.4177
	(8.8615)	(7.7927)	(8.8184)	(8.9557)	(9.4028)	(9.1017)	(8.5392)	(8.5540)
Years Since Migration	6.8542	5.7981	5.8920	9.0737	8.2659	8.8560	10.6829	8.5787
	(7.1428)	(6.5587)	(6.7178)	(7.6019)	(7.4536)	(7.6848)	(7.7642)	(7.3253)
Married	0.4408	0.3846	0.4458	0.4356	0.4045	0.4080	0.4390	0.4921
	(0.4966)	(0.4889)	(0.4972)	(0.4961)	(0.4917)	(0.4924)	(0.4973)	(0.5009)
Spouse Present	0.3242	0.2885	0.3060	0.3677	0.3521	0.3280	0.4024	0.3898
	(0.4681)	(0.4552)	(0.4609)	(0.4824)	(0.4785)	(0.4704)	(0.4914)	(0.4887)
Has Children	0.3108	0.2404	0.3163	0.3058	0.3146	0.2760	0.3008	0.3307
	(0.4629)	(0.4294)	(0.4651)	(0.4610)	(0.4652)	(0.4479)	(0.4595)	(0.4714)
Birthplace								
Italy	0.2174	0.0481	0.2625	0.1357	0.0524	0.1240	0.1301	0.2402
	(0.4126)	(0.2150)	(0.4401)	(0.3426)	(0.2233)	(0.3302)	(0.3371)	(0.4280)
Russian Empire	0.1783	0.3365	0.1004	0.3333	0.3970	0.3960	0.4146	0.1260
	(0.3829)	(0.4748)	(0.3005)	(0.4716)	(0.4902)	(0.4900)	(0.4937)	(0.3325)
Other new migrants	0.2037	0.2308	0.1523	0.3137	0.3483	0.2840	0.2967	0.3228
	(0.4028)	(0.4234)	(0.3594)	(0.4642)	(0.4773)	(0.4518)	(0.4578)	(0.4685)
Germany	0.1089	0.1346	0.1151	0.0924	0.0637	0.0600	0.0772	0.1693
	(0.3115)	(0.3430)	(0.3193)	(0.2898)	(0.2446)	(0.2380)	(0.2675)	(0.3757)
Ireland	0.0942	0.0385	0.1344	0.0118	0.0037	0.0280	0.0081	0.0079
	(0.2922)	(0.1932)	(0.3412)	(0.1080)	(0.0612)	(0.1653)	(0.0900)	(0.0886)
Other old migrants	0.1148	0.0962	0.1456	0.0492	0.0749	0.0480	0.0203	0.0512
	(0.3189)	(0.2962)	(0.3528)	(0.2163)	(0.2637)	(0.2142)	(0.1414)	(0.2208)
Other	0.0826	0.1154	0.0896	0.0639	0.0599	0.0600	0.0528	0.0827
	(0.2753)	(0.3210)	(0.2857)	(0.2447)	(0.2378)	(0.2380)	(0.2242)	(0.2759)
	Difference petition-declaration							
Log Occupational Score	0.0315	0.0174	0.0173	0.0640	0.0388	0.0510	0.0590	0.1083
	(0.5216)	(0.4327)	(0.4937)	(0.5847)	(0.6523)	(0.5833)	(0.6025)	(0.4862)
Age	5.2965	5.8824	5.3378	5.1460	5.3434	5.0738	5.0166	5.1317
	(2.1923)	(2.7622)	(2.1379)	(2.2324)	(2.6741)	(1.9252)	(2.1889)	(2.0224)
Years Since Migration	4.6982	5.0865	4.7325	4.5831	4.6779	4.4760	4.4472	4.7205
	(1.7170)	(1.6901)	(1.7086)	(1.7307)	(1.7921)	(1.6627)	(1.7436)	(1.7116)
Married	0.2177	0.3365	0.2191	0.2026	0.2060	0.2160	0.2195	0.1693
	(0.4128)	(0.4748)	(0.4137)	(0.4021)	(0.4052)	(0.4123)	(0.4148)	(0.3757)
Spouse Present	0.2228	0.3846	0.2267	0.1976	0.2022	0.2160	0.2033	0.1693
	(0.4439)	(0.4889)	(0.4437)	(0.4362)	(0.4382)	(0.4406)	(0.4419)	(0.4251)
Has Children	0.1721	0.2019	0.1747	0.1632	0.1573	0.2040	0.1585	0.1339
	(0.3775)	(0.4034)	(0.3798)	(0.3698)	(0.3648)	(0.4038)	(0.3660)	(0.3412)
N	3353	104	2232	1017	267	250	246	254

Standard deviations in parentheses. Distinctive refers to migrants for which $\Delta Ai < 0$; Keepers are migrants for which $\Delta Ai = 0$; Americanize refers to migrants for which $\Delta Ai > 0$; quartiles refer to migrants who Americanize.

Table 3: Effect of Name Americanization on Log-Occupational Score, Benchmark Specification

	OLS			First differences		
	I	II	III	IV	V	VI
A_i	0.0426**	0.0344*	0.0395**	0.1178***	0.1344***	0.1412***
	(0.0185)	(0.0190)	(0.0191)	(0.0376)	(0.0378)	(0.0385)
Marital status	0.0255	0.0346**	0.0355**	0.0332	0.0327	0.0228
	(0.0172)	(0.0174)	(0.0173)	(0.0269)	(0.0265)	(0.0272)
Has children	–0.0041	0.0029	–0.0032	0.0028	0.0027	0.0076
	(0.0127)	(0.0127)	(0.0127)	(0.0202)	(0.0209)	(0.0213)
Spouse present	0.0407***	0.0266*	0.0086	–0.0258	–0.0147	–0.0089
	(0.0142)	(0.0142)	(0.0147)	(0.0260)	(0.0265)	(0.0277)
Italy		–0.0936***	–0.0985***		0.0028	0.0086
		(0.0174)	(0.0195)		(0.0055)	(0.0065)
Russian Empire		–0.0336	–0.0445*		0.0061	0.0134
		(0.0209)	(0.0249)		(0.0067)	(0.0089)
Other new migrants		–0.0415**	–0.0392*		–0.0001	0.0055
		(0.0179)	(0.0205)		(0.0044)	(0.0054)
Germany		0.0366*	0.0306		–0.0028	0.0014
		(0.0198)	(0.0219)		(0.0048)	(0.0055)
Ireland		–0.0412*	–0.0419*		0.0186**	0.0219***
		(0.0220)	(0.0239)		(0.0075)	(0.0082)
Other		–0.0675**	–0.0622**		0.0355***	0.0424***
		(0.0277)	(0.0300)		(0.0111)	(0.0121)
$\Delta Trend_{it}$	Yes	Yes	Yes	Yes	Yes	Yes
Lab. Mkt. Ind.	No	No	Yes	No	No	Yes
R^2	0.01	0.01	0.06	0.01	0.01	0.03
N	6706	6706	6706	3353	3353	3353

Robust standard errors in parenthesis.
ΔA_i = change in the Americanization index A_i which is defined as 1 for American name with highest frequency and 0 as the American name with lowest frequency. Labor market indicators refer to dummy variables for each of the NYC community districts. See text for explanation.
All regressions include a time trend.
Country of birth and labor market indicators should be intended as interactions between these variables and the time trend.
Reference category for country of birth: Old migrants
Married, Has children, Spouse present in the U.S. are all indicators. See text for explanation.

Table 4: Effect of Name Americanization on Log-Occupational Score, Causality Tests

	Name changers			Instrumental variable			Instrumental variable		
	I	II	III	IV	V	VI	VII	VIII	IX
ΔA_i	0.1091***	0.1136***	0.1313***	0.8083**	0.5604*	0.6850**	0.8495***	0.5977**	0.7116**
	(0.0417)	(0.0424)	(0.0440)	(0.3497)	(0.3156)	(0.3458)	(0.3005)	(0.2727)	(0.2894)
N	1538	1538	1538	3353	3353	3353	3353	3353	3353
C. of birth Ind.	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Lab. Mk. Ind.	No	No	Yes	No	No	Yes	No	Yes	Yes
				First stage			First stage		
$S_{jArrival}$				0.0499***	0.0550***	0.0543***	0.1208***	0.1208***	0.1208***
				(0.0069)	(0.0072)	(0.0074)	(0.0105)	(0.0105)	(0.0105)
$S^2_{jArrival}$							–0.0416***	–0.0416***	–0.0416***
							(0.0071)	(0.0071)	(0.0071)
F 1^{st} stage				51.737	58.416	53.685	30.473	37.773	36.411
Partial R^2				.013	.016	.015	.018	.023	.023
Wooldridge test p-value				.038	.163	.103	.008	.074	.038
Sargan test							.859	.871	.915
N				3353	3353	3353	3353	3353	3353

Robust standard errors in parenthesis.
ΔA_i = change in the Americanization index A_i which is defined as 1 for American name with highest frequency and 0 as the American name with lowest frequency.
$S_{jArrival}$ refers to the scrabble index as explained in the text.
Wooldrdige test refers to a robust score test of endogeneity (Wooldridge, 1995).
All models include the covariates in the last column of Table 3.

Table 5: Effect of Name Americanization on Log-Occupational Score, Channels

	New Migrants				Old Migrants			
	OLS	FD	NC	IV	OLS	FD	NC	IV
ΔA_i	0.0445	0.0911	0.0988	–0.5149	0.0470*	0.1526***	0.1434***	0.6963**
	(0.0284)	(0.0566)	(0.0776)	(0.7953)	(0.0264)	(0.0467)	(0.0515)	(0.3171)
N	2070	1035	251	1035	4636	2318	1287	2318
	Internal Migrants				**Internal non-Migrants**			
	OLS	FD	NC	IV	OLS	FD	NC	IV
ΔA_i	0.0189	0.0559	0.0567	0.2148	0.0579*	0.2325***	0.1713**	1.2913**
	(0.0198)	(0.0353)	(0.0424)	(0.3116)	(0.0351)	(0.0743)	(0.0852)	(0.5298)
N	3830	1915	881	1915	2876	1438	657	1438
	High Exposure				**Low Exposure**			
	OLS	FD	NC	IV	OLS	FD	NC	IV
ΔA_i	0.0264	0.0081	0.0005	–0.0737	0.0366*	0.1550***	0.1595***	0.5619***
	(0.0375)	(0.0384)	(0.0444)	(0.3237)	(0.0204)	(0.0399)	(0.0484)	(0.2141)
N	1400	700	414	700	4202	2101	858	2101

Robust standard errors in parenthesis.
First Stage F-tests: 3.849, 33.232, 12.815, 17.67, 7.449, 25.537.
ΔA_i = change in the Americanization index A_i which is defined as 1 for American name with highest frequency and 0 as the American name with lowest frequency. See text for explanation. All models include the covariates in the last column of Table 3

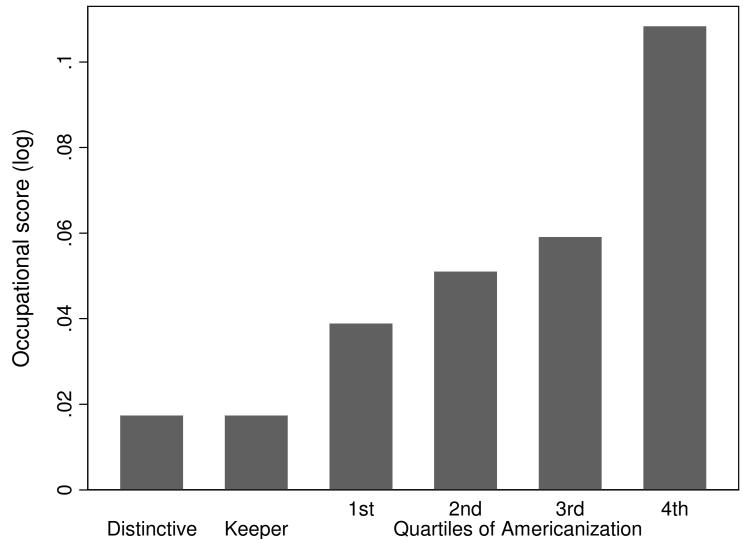
Table 6: Effect of Name Americanization on Log-Occupational Score, Robustness checks

	Type 3				Early Declarants			
	OLS	FD	NC	IV	OLS	FD	NC	IV
ΔA_i	0.0314	0.2385**	0.2162	5.0440	0.0345	0.1017**	0.1058*	0.9124*
	(0.0296)	(0.1208)	(0.1506)	(5.5015)	(0.0267)	(0.0485)	(0.0592)	(0.4771)
N	4374	2187	372	2187	3042	1521	571	1521
	Soundex				**All 1930 Names**			
	OLS	FD	NC	IV	OLS	FD	NC	IV
ΔA_i	0.0397**	0.1280***	0.1125**	0.9793**	0.0521***	0.1432***	0.1329***	0.6156**
	(0.0178)	(0.0412)	(0.0444)	(0.4463)	(0.0199)	(0.0406)	(0.0467)	(0.2468)
N	6706	3353	1538	3353	6706	3353	1538	3353

Robust standard errors in parenthesis.
First Stage F-tests: .766, 8.599, 22.041, 49.729, 35.478, 25.776, 32.958, 34.675.
ΔA_i = change in the Americanization index A_i which is defined as 1 for American name with highest frequency and 0 as the American name with lowest frequency. See text for explanation. All models include the covariates in the last column of Table 3

9 Figures

Figure 1: Change in Occupational score and Americanization



The x-axis represents quintiles of the change in the Americanization. The category 0 indicates individuals who did not change name. The dependent variable is the occupational score in 1950 U.S. dollars multiplied by 100.

Figure 2: Effect of Changing Name into a Popular American Name

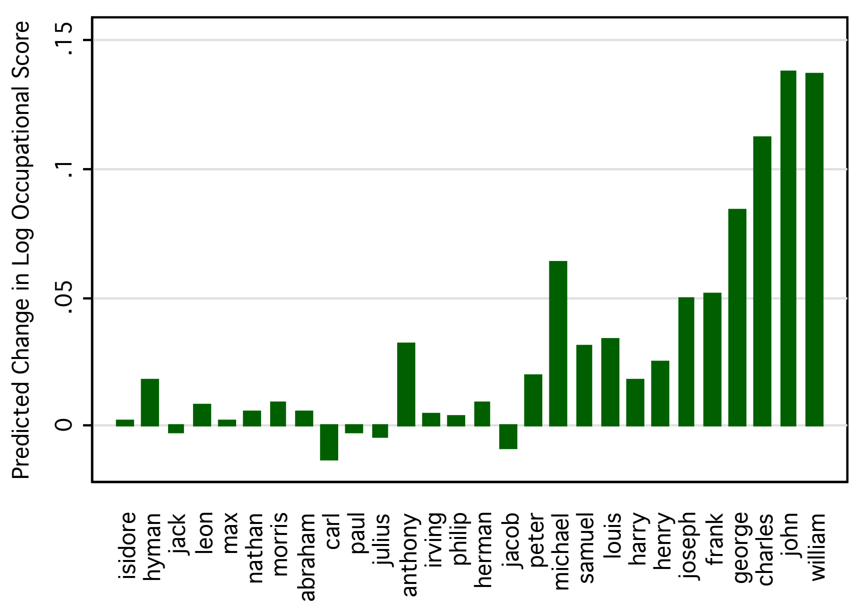
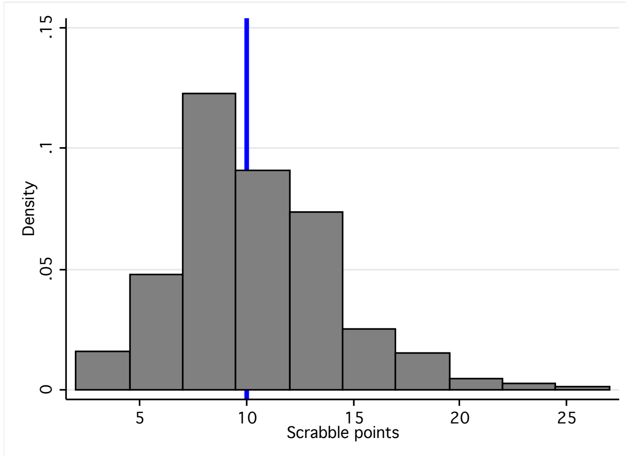
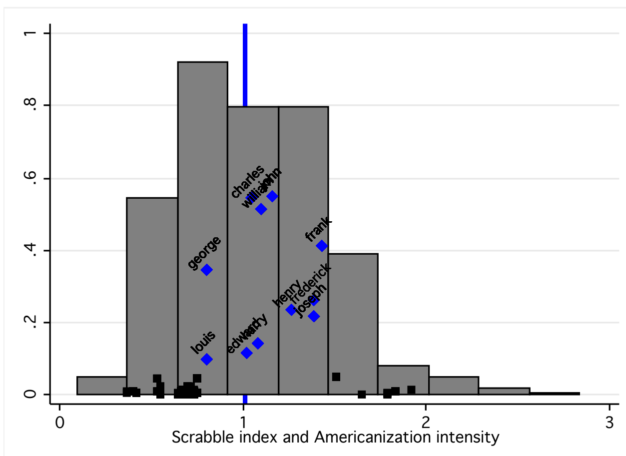


Figure 3: Scrabble points, Scrabble index and Americanization Intensity

(a) Scrabble Points, U.S. born population 1880



(b) Scrabble index and Americanization Intensity



(c) Difference in Scrabble index between petition and arrival

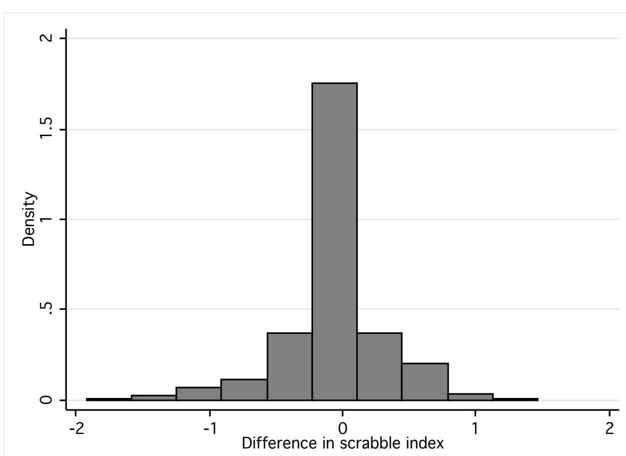
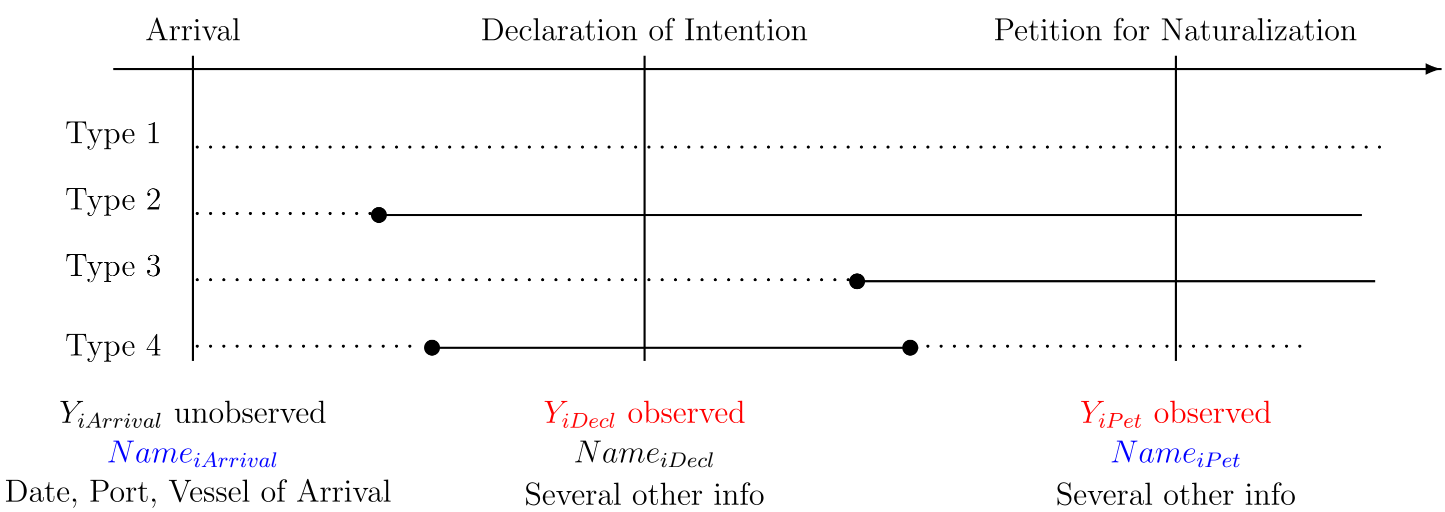


Figure 4: Timing of name changes and naturalization documents



References

Abramitzky, R., L. Boustan, and K. Eriksson (2012a). A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. Working Paper 18011, NBER.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2012b). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review 102*(5), 1832–1856.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Arai, M. and P. S. Thoursie (2009). Renouncing personal names: An empirical examination of surname change and earnings. *Journal of Labor Economics 27*(1), 127–147.

Aura, S. and G. D. Hess (2010). What's in a name? *Economic Inquiry 48*(1), 214–227.

Bartel, A. P. (1989). Where do the new us immigrants live? *Journal of Labor Economics*, 371–391.

Bertrand, M. and S. Mullainathan (2004). Are Emily and Greg more employable than Lakisha and Jamal? a field experiment on labor market discrimination. *The American Economic Review 94*(4), 991–1013.

Broom, L., H. P. Beem, and V. Harris (1955). Characteristics of 1,107 petitioners for change of name. *American Sociological Review 20*(1), 33–39.

Calvó-Armengol, A. and M. O. Jackson (2002). Social networks in determining employment and wages: patterns, dynamics and inequality. *American Economic Review*.

Calvó-Armengol, A. and M. O. Jackson (2007, January). Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory 132*(1), 27–46.

Drury, D. W. and J. D. McCarthy (1980). The social psychology of name change: Reflections on a serendipitous discovery. *Social Psychology Quarterly*, 310–320.

Ferrie, J. P. (1996). A new sample of males linked from the public use microdata sample of the 1850 us federal census of population to the 1860 us federal census manuscript schedules. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 29*(4), 141–156.

Ferrie, J. P. (1999). *Yankeys Now: Immigrants in the Antebellum U.S. 1840-1860*. Oxford University Press.

Figlio, D. N. (2005). Names, expectations and the black-white test score gap. Technical report, National Bureau of Economic Research.

Fryer, R. G. and S. D. Levitt (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics 119*(3), 767–805.

Goldin, C. and M. Shim (2004). Making a name: Women's surnames at marriage and beyond. *The Journal of Economic Perspectives 18*(2), 143–160.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 1360–1380.

Hatton, T. J. and J. G. Williamson (2004). International Migration in the Long Run: Positive Selection, Negative Selection and Policy. Discussion Paper 1304, Institute for the Study of Labor.

Higham, J. (2002). *Strangers in the land: Patterns of American nativism, 1860-1925*. Rutgers University Press.

Kang, T. S. (1971). Name change and acculturation: Chinese students on an american campus. *The Pacific Sociological Review 14*(4), 403–412.

Laham, S. M., P. Koval, and A. L. Alter (2011). The name-pronunciation effect: Why people like Mr. Smith more than Mr. Colquhoun. *Journal of Experimental Social Psychology*.

Lleras-Muney, A. and A. Shertzer (2012). Did the americanization movement succeed? an evaluation of the effect of english-only and compulsory schools laws on immigrants' education. Technical report, National Bureau of Economic Research.

Long, J. and J. Ferrie (2013). Intergenerational occupational mobility in great britain and the united states since 1850. *American Economic Review 103*(4), 1109–37.

Steven Ruggles, J., A. Trent, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010). Integrated public use microdata series: Version 5.0 [machine-readable database]. Technical report, Minneapolis: University of Minnesota.

Wooldridge, J. M. (1995). Score diagnostics for linear models estimated by two stage least squares. In G. S. Maddala, P. C. B. Phillips, and T. N. Srinivasan (Eds.), *Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao*, Chapter 3, pp. 66–87. Oxford: Blackwell.